DOCUMENT RESUME

ED 237 121 JC 830 443

AUTHOR Hartman, Neal E.
TITLE Maximizing the Effectiveness of Reading Tests in the Community College.
INSTITUTION Saint Louis Community Coll., Mo.
PUB DATE Sep 81
NOTE 55p.
PUB TYPE Reports - Research/Technical (143) -- Statistical Data (110)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Academic Achievement; Community Colleges; Cutting Scores; Educational Counseling; Feedback; *Reading Tests; *Student Placement; *Test Validity; Two Year Colleges

ABSTRACT

A study was conducted at St. Louis Community College to validate the Nelson-Denny Reading Test (NDRT), Form C, as an instrument for student placement, to empirically establish a cutoff score for a mandatory reading improvement program, and to provide more complete and meaningful feedback to students about their reading test results. The study sample was composed of day students with less than one semester's college credit and with a work plus study load not exceeding 72 hours per week. The initial sample of 500 students tested was reduced to 186 as a result of the imposition of various controls. Results of a three-phase analysis of NDRT scores and course grades included the following: (1) the mean vocabulary, comprehension, and total NDRT scores of the sample fell into the 31st, 32nd, and 28th percentiles, respectively, according to national norms; (2) the mean grade of the sample, adjusted for withdrawals, was 2.130; (3) reading skill level accounted for almost one-third of academic achievement variance; (4) 74% of those who scored above 49 on the NDRT passed, 28% who scored above 49 failed, and 29% who scored below 49 passed; and (5) with a 49 on the NDRT as a cutoff, 34% of the incoming students would qualify for reading improvement courses. A final phase of the study involved generating feedback data, including national and local percentiles and a predicted GPA (grade point average), for use in counseling and advisement. (HB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED237121

MAXIMIZING THE EFFECTIVENESS OF READING TESTS IN THE COMMUNITY COLLEGE

Neal E. Hartman
St. Louis Community College
St. Louis, Missouri

September, 1981

©

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

N. E. Hartman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

X Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

VC 830 443

CONTENTS

Preface 1

Introduction 3
Phase I - Validation 8
Phase II - Cutoff Development 13
Phase III - Feedback for Counseling and Advising 15

Method 16
The Test 16
The Controls 17
Sample Size 22
The Analyses 22

Results 25
Table 1. Descriptive Statistics 26
Figure 1. Vocabulary Raw Scores 27
Figure 2. Comprehension Raw Scores 28
Figure 3. Total (V+C) Raw Scores 29
Figure 4. Cumulative Percent Ogive - T Scores 30

Phase I - Validation 33

Table 2. Coefficients of Correlation Test Scores and spa 33

Phase II - Reading Requirements Cutoff 35

Table 3. NDRT Cutoff Analysis 36
Table 4. Cost-Effectiveness Rates for the Lowest, Optimum, Highest Cutoffs 40

Phase III - The Feedback 40

Summary and Discussion 41

References 48

Appendix Norms and Predicted gpa's 49

Preface

This report is the culmination of the efforts of many people. It grew, out of discussions in the Reading Task Force, established by Dean Duvall and chaired by Ms. Anita Taylor. The task force members were Ken Boyer, Jane Bracy, Bob Martin, Joan Mizis, Anita Taylor, Benjamin Young, and the author. Each member above contributed significantly in establishing a need for a study and in determining the direction it should go. Anita Taylor was especially instrumental in initiating the original study, of which this one is a follow-up.

The reading tests were administered and scored by the Reading Department, under the direction and coordination of Marge Bagwell, who also assisted with recording test data cards, which she also assisted in sorting.

A note of thanks is also due the instructors in whose classes the test was administered: Marvin Barnett, Grace Tripp, Robert Martin, Rolland Haun, Robert Murphy, Daniel Conoyer, Mary Petropoulos, Marge Park, and Stan Kary.

Acknowledgement is particularly due the Florissant Valley counseling staff who permitted the writer additional time from other counseling duties to conduct the research and to write this report. In a real sense, the major effort of the data sorting, data analysis, and report writing was a combined effort of the Counseling staff. The Director of Counseling, Benjamin Young, was especially facililative in providing the time and the flexibility to do the work, and without whose cooperation and encouragement, this study would not have been completed.

Although all counselors participated, either directly or indirectly, the following are owed a special debt of gratitude for their assistance.

To Lee West for her assistance in data recording and especially for her suggestions on how to determine "full load equivalents" for course load and job load.

To Pat McCracken for his individual suggestions on how to analyze the significance of the total withdrawal data.

To Lou Guese, and Vic Hoemann for their patient assistance in checking the accuracy of data on yards of calculation tape.

To Barbara Hunt for her invaluable help in editing the rough drafts, and for helping to correct the grammar and spelling errors, any of which remaining are the sole responsibility of the writer.

To Mrs. Lorraine Lubker and Mrs. Debra Maxey for typing and retyping several drafts, often from handwritten manuscripts which were not paragons of penmanship.

To all of the above, and to yet others not named, much credit is due. For any weaknesses of the research design or of implementation, the writer takes full responsibility.

Neal E. Hartman

INTRODUCTION

The general purpose of this study was to investigate the broad, pervasive relationship that is presumed to exist between reading achievement level and academic success for college freshman. This was to be accomplished by examining in detail the relationship of reading test scores to measures of academic achievement. The investigation was done in three separate parts, or phases. It is generally assumed by educators and it is often observed casually, that there is definitely a relationship between reading ability, or achievement, and academic success. That undoubtedly is the reason for the plethora of remedial reading courses which have been instituted in college and universities. But there has been relatively little empirical investigation as to what level of reading achievement is necessary for specifically defined, or quantified, levels of academic success. In most discussions of academic success as a function of reading ability the terms academic success and reading ability (achievement) have an assumed common meaning to the discussants, or those terms are, at best, poorly defined. Thus, these terms, as used in this study shall be defined operationally later in this paper.

A basic question is how well should a beginning college student be able to read to succeed? It is no longer sufficient merely to state that too many of our students are reading at the fifth or eighth grade level, or that they are "functionally illiterate" (whatever that means) and that "something" should be done about this deplorable situation. Something indeed must be done, but a specific target for improvement needs somehow to be established

rather than just using the shotgun approach of assigning "poor" readers into reading improvement classes, often against their will and without having any definite outcome of the classes except "improvement."

Thus, establishing a more precise level of reading achievement that would be the minimum to have a reasonable expectation of college success was a specific purpose of this study. A corollary purpose was to help establish a methodology which may be used in any educational setting and which may be used repeatedly as the relationship between reading and academic success changes over time. In order to accomplish these goals, the research was conducted in three phases:

> Phase I (validity) was to determine statistically the overall relationship between reading test scores and general academic achievement; in essence, this phase was to validate the reading test used for a specific population as an instrument to place students into reading classes. As shall be seen below, the degree of validity of reading tests used in colleges is mainly intuitive, rarely quantified.
>
> Phase II (cutoff) was to determine empirically the relationship of various different reading test scores to the minimum academic achievement level that could be considered "academic success." This phase, then, was to establish the most efficient reading test score as a cutoff for requiring students to enroll in a non-credit reading improvement course. This phase, of necessity, was conducted only after the degree of test validity was determined.
>
> Phase III (feedback) was to determine the relationship of any individual's reading test score to the scores of other students just commencing their college programs. In short, local norms were established and in addition, a predicted first semester gpa was established for each possible reading test score.

Various reading test reviewers and reading test manuals have recommended that studies of this type be done; few, if any, provide actual examples of such studies having been done, nor do they indicate very clearly how such studies should, or could, be done. They are simply "well advised."

Such studies are more than well advised. They are essential to any well thought-out program of reading improvement which would be administered fairly (non-punitively) and which would produce results clearly meaningful. That statement refers to the fact that with the increased emphasis on universal access to higher education there came a surge of students the Chicago Tribune described as "severely unprepared for college work" (O'Connor, 1980), and that these students are being required to engage in programs of reading improvement before they can enter the educational and training programs of their choice.

Students who are to be prohibited from exercising a freedom of choice must be provided with valid evidence as to why they should be "required" to do whatever, and that evidence should be empirically derived rather than being arbitrary, capricious, or subjective. Otherwise, the students will feel discriminated against, will resent the requirement, and will, therefore, not be motivated to take advantage of the opportunity for self-improvement.

It is understandable that reading programs should be required and continued for those with little, or marginal, reading skills. The chancellor of Chicago's City Colleges, Dr. Oscar Shabat, was quoted as saying that if you have them (students) and you're mowing them down because you believe they have a background they don't and you treat them accordingly, the result is that there is tremendous failure (O'Connor, 1980).

This "tremendous failure" is not exclusively a Chicago problem. It has become a national problem and exists wherever there are colleges with open doors through which "new students" are encouraged to enter (Cross, 1976). "These students' failure is our failure," states Chancellor Shabat. In order to rid ourselves of this failure syndrome there now is a demand not only for

equal access to higher education, but also a demand for an equal opportunity to learn once access is attained (Cross, 1976). Thus, as a response to these new demands, most open admissions colleges and universities have established extensive developmental and remedial learning programs, some of which include elaborate learning skills centers. Obviously one of the most important skills taught is reading.

Reading skill, or achievement, as measured by a reading achievement test, has long been seen as one of the academic skills most necessary to possess in order for one to achieve in college; however, far too few persons have adequate reading skill and far too many students, especially the "non-traditional" ones, may be characterized as being functionally illiterate: "An overwhelming number of the 63,000 students in the eight two-year colleges (in Chicago) are severely unprepared for college work. The average student reads at the 5th grade level" (O'Connor, 1980). It could scarcely be argued that college students reading at the 5th grade level are not illiterate.

Yet, according to Rouche, (1978), a person is not functionally illiterate unless he/she reads below the 4th grade level, a seemingly arbitrary definition; however, another definition of "illiterate" is supplied by Cohen (1976) which may be more applicable. It is relevant to any situation: "Functional literacy is that level of reading, writing, and calculating ability which a person needs to succeed in the public realm in which he chooses to operate."

Therefore, by applying Cohen's relevant, or situational, definition of functional illiteracy to Rouche's finding that the average community college student reads between the 8th and 9th grade level, it follows that the "average" community college student is functionally illiterate - given the assumption that the texts and the instruction are at the college level. From this per-

spective, it means that approximately half of the students in community colleges are functionally illiterate, and if they are indeed to have the freedom to learn, (as well as the freedom to fail) steps must be taken immediately to raise the level of literacy to at least a minimal point of functional literacy.

Thus, around the country there are many courses, or programs, of basic reading skills improvement based upon competency. Chancellor Shabat has indicated, "Every student will be given a reading test. We're going to be placing our students on the basis of reading ability...Really poor readers, even those with high school diplomas, may be barred from the colleges and placed instead in the Urban Skills Institute, a pre-college program" (O'Connor, 1980). For similar reasons, various curricula at many open admissions colleges now may be entered only after careful academic screening. It has further been suggested that placement be done on a course-by-course basis (Cohen, 1975); however, the logistical problems attendant to that suggestion could be staggering.

As the practice spreads of requiring (not merely offering) reading improvement courses, or programs, based on reading test scores, it may be reasonably questioned whether or not the reading tests are valid, and below what score could students be required to forego a course, or courses, of their own choice and take instead reading improvement? These two important aspects of a reading program, reading test validity and the establishment of a fair cutoff score, were only inferred by Chicago's Shabat by his stating that the first step is an "honest and accurate assessment of student skill through reading tests ... The second step is through research to determine the student success rates," (necessary to

determine accurately a cutoff score) "as well as research into effective and innovative methods of teaching." Accordingly, the research reported here has addressed these two basic concerns of test validity and student success rates.

The "student success rate" concern has been addressed here in two phases: the cutoff phase and the student feedback phase (Phases II and III). Phase I, of course, was of necessity, the test validation phase. A brief statement of the purpose of each phase is provided below; more details of how these phases were studied are provided later in the methods section.

Phase I - Validation

Ideally, in any reading program based upon reading test scores, the validity of the reading test should be determined for the population which will be tested. At the very least, validity should be noted from research conducted at institutions with very similar populations. Unfortunately, very few institutions validate their own tests; it is difficult to find any institution with a similar population to the one used here which has conducted any type of rigorous validation study. But any well conceived reading program should do its own validation study. In addition to using its own population to validate the test, the criterion against which the test is to be validated should be relevant to the intended use of the test - whether for screening, diagnosis, prediction, or whatever.

In most instances, including this one, the primary determination must be whether or not the test actually does discriminate between the better and the poorer readers. Of course that is of primary interest because of the usually implicit assumption that underlies all reading programs, and all

reading testing programs: It is necessary to read well in order to do well academically; better readers achieve higher academically than poorer readers; there is a reading level below which students cannot expect to achieve sufficiently well to justify the expense in time and money for them to be in a college setting (they are functionally illiterate relative to that setting). Given that this overall assumption is correct, a test which truly does discriminate between better and poorer readers would also discriminate between better and poorer academic achievers. It should follow, then, that the degree of validity of the test may be established by determining the strength of the relationship between test scores and academic achievement. Or, academic achievement would be the criterion against which the test would logically be validated. Such was the case in this study.

The statement of the above assumptions may appear to be a needless exercise in simplistic tautology, but all too often we as educators lost sight of these fundamental relationships. It is hoped that by making these assumptions more explicit, a better understanding of what a reading test can do, as well as its limitations, will result.

Finding a suitable reading test to use at the college level for a comprehensive reading program is no small task. A search of Buro's (1978) Mental Measurements Yearbook (MMY), revealed a lack of such tests. Only two appeared to be appropriate. One was the Nelson-Denny Reading Test (NDRT). The other was the Iowa Silent Reading Test (ISRT).

Of the ISRT it was stated, "It is difficult envisioning the ISRT - or its competitors - as having much diagnostic ability. For assessment purposes, however, the 1973 ISRT series is, all things considered, an impressive

addition to the current inventory of reading tests" (Hakstian, 1978). But in the same edition of Buros' MMY Frederick Davis (1978) states that the ISRT should be used as a special and intensive reading rest, not as a part of a comprehensive reading program. Practical considerations also limit the use of the ISRT in that it takes an inordinate amount of time to administer - 96 minutes in two session for level 3, grade 11-16. The ISRT's chief competitor, The Nelson-Denny Reading Test (NDRT), by comparison takes only 35 minutes in one session to administer.

From a practical standpoint, therefore, the NDRT is probably the better of the two tests, and was the one used here.

In his review of the NDRT, Alton A Raygor (1978), psychology professor and coordinator of the Reading and Study Skills Center at the University of Minnesota, pointed out that the college norms for the Nelson-Denny are a mixture of two-year and four-year schools: "No effort is made to differentiate between the two types of schools in spite of widespread evidence of rather large differences in measured reading abilities." The NDRT manual confirms that the norms are, in fact, derived from an amalgam of eleven four-year colleges and six two-year colleges. Given the previosuly stated finding, that approximately half of all community college students are reading at or below the 9th grade level and are, therefore, functionally illiterate (Rouche, Cohen), Raygor's concern about the mixed types of colleges the norms represent is a legitimate concern. Such a mixture makes score interpretation more confusing than if specific normative data were available for specific types of colleges. The amalgamated norms published in the NDRT manual are not representative of either a four-year college or a two-year

college, and because there was no success in obtaining a stratified sample of students from the various types of colleges nationally, the norms do not really represent a national sample either. Thus, it would seem tnat the published "norms" are really normative of nothing and have only ordinal meaning, and acquire further meaning only as the test is used for a long time by an institution so that an intuitive validity develops within that institution. Forsyth (1978) reports in the MMY, "The normative data supplied for grades 13 through 16 do not permit any norm referenced interpretations of the test scores. The college level norms cannot be considered as repreenting any well defined population...At the college level it may have some validity for identifying people who need remedial work. However, little empirical evidence is given to support the use of the test as a screening device in any setting...Actually, the given data represent rather weak evidence of predictive validity" (p. 1208, emphasis added).

Again, then, one of the major purposes of this study was to determine for a specific institution if there is predictive validity and to quantify it, and furthermore determine how, if the NDRT is valid for this population, it may be utilized as a screening device.

Not only did the NDRT reviewers find the normative data to be almost meaningless and the predictive validity to be vague and ambiguous, but they also found the test to be rather lacking in face validity. That is, it did not seem to be written at an appropriate level to be used to screen college freshmen who might possibly be considered to be functionaly illiterate, relative to the college setting: "To use this test on anyone other than highly selected students will produce very questionable results. Worse than that, it will provide a difficult, defeating, and frustrating experience to

many of the students - especially those who have already had enough of such experience" (Raygor, 1978). Raygor did not appear to exagerate. Note the content of the first essay encountered by the student, the first sentences of which, are quoted below:

> Thespis was supposed to have been the first poet who stepped out of the chorus and devised a dialogue with the members to make his poem more vivid. He was as Iscarian, his first official performance is supposed to have taken place in 534 B.C.

Continuing on, using complex grammar, the essay discussed in much detail the towns in Ancient Greece of Athens and Eleusis. If further discusses the poet Aeschylus, and the characters of Agamamnon, Clymnestra, and Orestes; it also refers to the gods Prometheus and Zeus. Such fare in the age of instant gratification as provided by television would scarcely pique the interest of the "traditional" student, much less the non-traditional student some of whom might not have graduated from high school. Of course the reasons for such passages are understandable and would be legitimate for use by the so-called traditional collegesand college students in the day when most college students followed a traditional "college prep" program in high school and mainly those in the upper half, or higher, of their graduating classes were accepted for college.

In short, the MMY reviews seem to indicate that there are no really adequate reading tests for screening and predicting at the college freshman level, especially for the open admissions colleges. In spite of that, one of the reviewere, Forsyth (1978), summarizes his review by stating "The Nelson-Denny Reading Test appears to be most useful at the college level. At this level, if local norms are developed (emphasis added), the test may validly serve the predictive, screening, and diagnostic purposes for which

it was developed" (p. 1208). That is precisely what the first phase attempted to accomplish.

Phase II - The Cutoff Norm Development

The use of a reading test to determine who will or who will not be required to enroll in a certain remedial program obviously infers that there has already been determined, in some manner, a cutoff score, at or below which students are required to enter into the remedial course, or courses. Thus, once a test has been selected and locally validated, the next step should be to determine empirically a cutoff score. Only after this process has occurred can there be realistic confidence that the best (or closest) cutoff possible is being used.

A cutoff may be established either intuitively or empirically. If purely chance factors alone both methods might result in an identical cutoff score. Even if that were the case, the intuitive method can provide no information as to the penalty or cost factor of the cutoff score; that is a major advantage of the empirical method. A cutoff set too high can result in very few people failing to succeed academically, but can also cause more students than necessary to take the required non-credit remedial course(s). This "cost" factor in cutoff determination is discussed in more detail in the results section.

Although the NDRT was never satisfactorily validated for any specified population, as reported by the manual it does recommend that a cutoff score be established when the test is to be used for screening purposes. "The establishment of an appropriate cutoff point...can help to distinguish students

who would benefit from placement in a developmental or in a remedial reading program"(NDRT Manual, p. 14).

The manual describes how, over a five year period, a cutoff was established - or discovered - at the University of Minnesota. Freshmen were administered the NDRT at the beginning of their college careers. Four years later, after they were graduated, a cutoff was determined. It should be noted that by the time the analysis was made there was at least a five year time lag, so that any cutoff would tend to be outdated. The Manual describes their method:

> In the Nelson-Denny Reading Test as entering freshmen, those graduating in four years with the distinction 'average' at the 42nd percentile, and in reading those graduating with distinction at the 58th percentile, and those graduating with distinction at the 42nd percentile. (NDRT Examiner's Manual, p. 14)

In the above example, it is not clear what was meant by 'average' - median, mode, or mean? Also it is not clear as to what they determined to be their cutoff. Presumably it was the 42nd percentile (on the "amalgamated" norms stated in the Manual or on local Minnesota norms?). Furthermore, there is no information as to the efficiency of the 42nd percentile, or whatever they decided to use as a cutoff. That is, what portion of students below the cutoff would be expected to "succeed" regardless of being below the cutoff, and what portion would not "succeed" even though they were above the cutoff? That such questions are frequently not addressed in the use of tests in education is probably one major reason that test usage has come under heavy attack in recent years (Nader, Nairn, and others, 1980).

This study has attempted to address itself to these and other vital concerns. And, although the NDRT manual did state that an informal study as the

one they cited should be followed up by a more careful statistical analysis, it would seem that it is putting the cart before the horse; also, they did not suggest any methodology as to how to conduct a "careful statistical analysis." Again, that is one of the purposes of the present study.

Phase III – Feedback for Counseling and Advising

Janice Abel (1980) states in "Academic Advising: Goals and a Delivery System" that the advising program must provide opportunities to students to identify their special learning needs. This should include information on a range of special characteristics (e.g. reading skills) which identify the intellectually gifted, the handicapped, ... or the emotionally troubled. Clearly, this includes everyone," she states. This is in fact another purpose of this study is to provide reading with feedback information to the total range of students who enter this community college. It is not sufficient to merely inform students that "yes you do" or "no you don't" have to take remedial reading work. Whenever students are required to take a test, they should ethically be given feedback that is both complete and accurate.

More complete feedback would also include comparisons of each student with the total group of their peers both locally and nationally. That was done by providing local percentile norms displayed along with the "national" ones provided by the Manual. Additionally, success point averages (spa's) were predicted and placed into tabular form for each total raw score on the NDRT, based, again, on the local population.(The spa is an adjusted gpa and gives a better indication of actual academic success.) Thus, the local percentile norms, side-by-side the "national" norms, plus predicted spa's

for the semester of entrance, plus a knowledge of the efficiency of the established cutoff should maximize greatly the quantity and the quality of the feedback given to students in counseling and advising.

This study, then, had three broad purposes: To validate the test, to empirically establish a cutoff score for a mandatory reading improvement program, and to provide more complete and meaningful feedback to all students about the results of their reading tests.

METHOD

This study utilized what might be termed as the "ex post facto" method. That is, it was conducted during the course of the regular, on-going process that students go through in obtaining an education, rather than using an experimental model with laboratory, replete with experimental and control groups. In essence, the institution itself was the laboratory, the "treatment" was attendance as a first semester college student, the independent variable, as it were, was reading ability level, as measured by a reading test. Academic success, as measured by a modified grade point average called a "success point average," or spa, was the dependent variable. The relationship between test scores and spa's was analyzed from three perspectives, or phases - phases I, II, and III. Prior to the three phase analysis, however, a brief status study was conducted to determine the means, and variances of the major variables in order to provide a reference point from which to better comprehend the subsequent three phase analysis.

The Test

The Nelson-Denny Reading Test (NDRT), Form C, was the measuring instrument used to indicate reading achievement. The reading department normally uses

Form D, but for the purposes of this study, the reading department recommended that the parallel form (Form C) be used. The NDRT manual provides ample evidence that the two forms are indeed closely parallel.

The Controls: Variables for which limited control was exercised

In an in vivo study, it is not feasible to employ the classical research model utilizing experimental and control groups. Thus, the variables that could be controlled to some degree were controlled, mainly in the process of sample selection. It should be obvious that in this type of research a multitude of variables simply cannot be accounted for; however, a few can be controlled to a limited extent. In this study ten of the more critical variables were recognized and accounted for. These are discussed below.

Prior college experience. The amount of prior experience in college was a major variable which could be controlled. It was assumed that experienced students, who had survived one or more semesters of college, if they had identical reading test scores as new students, should be able to achieve at a higher academic level than the new students - those with less than a full semester of college credit. Thus, all students who were in the initial sample, who had previously attempted 12 college credit hours or more, were deleted from the study because 12 credit hours is the minimum that is considered to be full-time. Conversely, those with 11 or fewer hours were considered "new," and were retained; therefore, the sample was composed of students with less than one semester's credit - less than 12 semester hours.

Overall job-study load. The total weekly work load of job hours added to credit hour load was another variable that was controlled within limits. It was assumed that students who had too heavy a load, (job hours plus credit

hours) would have a reduced probability of academic success due to the overload. That is, variance in academic achievement would be unduly affected by overall load and would contaminate the effect of reading achievement - the independent variable. More or less the opposite was assumed to be the case with students who took less than a full-time load (full time being 12 semester hours or more) and worked less than a full-time job (35-40 hours). Thus, all students retained in the sample must have had total work week loads within a "normal range." That meant work plus study loads approximately equivalent to students enrolled in from 12 to 18 credit hours (the full-time load limits) with either no job or only a part-time job. The traditional rule of thumb of two hours of homework for each credit hour enrolled in was used to arrive at a "formula" which arbitrarily had as its lower limit 36 hours of combined job-classroom-homework and 72 hours as the upper limit. It was also stipulated that a student at the lower limit must be in at least six credit hours of coursework; thus, for example, a student enrolled in only a three hour course and working 35 hours on a job would have been excluded for having too light a load.

Age. Age was controlled for informally by virtue of the fact that only daytime students were tested. Most full-time day students are in the 18-19 year age bracket. Therefore, because age was relatively homogeneous and prior college experience was also, it could be inferred that overall life experience (which affects learning and motivation) also was homogeneous. In this instance, that was a desirable sampling attribute so that variance in academic achievement could better be attributed to variance in reading achievement level rather than to life experience, or maturity (age).

Sex. Sex differences were deliberately not accounted for. The sex composition of the sample was quite representative of the institution's population. The major reason that sex differences were not controlled for here was so that if significant results were found they could be applied across the board by the institution. That is, for example, a cutoff score for requiring reading improvement would be much easier to administer if it were the same for both men and women than to use a separate cutoff for each sex. The final sample, then, contained 89 males and 97 females, which closely approximates the M:F ratio of the campus.

Reading improvement during the semester. Reading improvement during the course of the first college semester was controlled by excluding from the analysis students who were then currently enrolled in a reading/study skills course. The assumption, of course, was that such courses do in fact improve reading and study efficiency. If that assumption were valid, the inclusion of such students would have largely nullified the meaning of any relationships that might have appeared between entering student reading test scores and first semester academic achievement.

Withdrawals. Withdrawals from courses were of two types. Each type required a special type of control. There were (a) total withdrawals and (b) partial withdrawals.

Students with total withdrawals were totally deleted from the study. Total withdrawals obviously would have 0.000 grade point averages; to use 0.000 gpa's as a criterion, the assumption would have to be made that no (zero) learning in any of the courses occurred and a further hypothesis would be made, implicitly, that the reading level of the students who totally

withdrew was below a threshhold of measurement and also below the threshhold of college learning. Those assumptions would have been necessary in order to use numbers as a ratio or equal interval scale as was assumed by the other statistical analyses of the study. Such assumptions, of course, could not be warranted. Instead of reading level being below a hypothetical threshhold of measurement or learning being the cause of total withdrawal, other factors such as job changes, personal/financial troubles and similar reasons were assumed to be contributing significant causes of withdrawal. Such a multitude of factors could not be controlled for; furthermore, to check out the rather informal hypothesis that there was no difference in reading skill level between the total withdrawal group and the group which did not totally withdraw, a t-test using the pooled variance formula was done to determine if the difference between the reading scores of the two groups differed significantly. A t score of 0.506 indicated the hypothesis of no difference must be accepted, i.e. the assumption that the total withdrawees, as a group, had essentially the same reading skills as the remaining group which included only partial withdrawals, was validated; therefore, greatly reduced reading skill was not the major reason for total withdrawal.

Students who had partial withdrawals were retained in the study but their W's were recomputed into their grade point averages to reflect the fact that for the courses withdrawn from no (zero) credit was earned. In essence, those withdrawals were treated as "F" grades are. The present grading system ignores partial withdrawals in computing the grade point averages. The recomputed gpa's were termed success point averages (spa's). As was the case for total withdrawals, it is conceded that outside factors in the life of the student could also be a reason to drop a course, or courses; however, most students

have to overcome some minor problems and the implicit assumption underlying this type of study is that students with better reading skills can cope with those problems better and, as a group, achieve at a higher level academically. That assumption is part of what this study is all about; thus, the partial withdrawees were retained with grade point averages being converted to success point averages (spa's).

Instructor variability. This was another variable that produces variability in the criterion - academic performance. This variable was controlled for informally. First, it was decided to utilize as a criterion all the grades a student received rather than just using the grade from one single class (as was done in a prior analysis), (Hartman & Bagwell, 1980). Thus, grades from instructors who were "easy graders" and grades from "hard graders" would tend to balance out. Also the different intrinsic interest value of one course over another course across the sampling of students would also balance out.

That the students were predominately full-time day students meant that few, if any, had part-time instructors who may or may not have similar teaching-grading philosophies and practices as their full-time counterparts. Thus, quality, experience, and dedication of instructors was maximized and held approximately equal for the students involved.

Native language. English as a native language was easily accounted for by sample selection. Any student whose native language was not English was deleted from the sample for obvious reasons.

Handicapped students. Students with severe handicaps which might impede academic performance were not retained in the study.

Knowledge of reading test scores. Knowledge of the students' reading test scores by either the students or the instructors could conceivably affect academic performance. Thus, the reading test scores were not revealed to instructors or students until the end of the semester.

In order to screen on many of the above variables, such as sex, age, class load, etc., the students filled out a questionnaire at the time they were tested.

Sample Size

The gross sample size which was initially tested and given the questionnaire was approximately 500. After sorting to exercise the controls, the net beginning of the semester sample was 201. The end of the semester net sample was further reduced to 186 as a result of the 15 students who withdrew totally during the course of the semester. It might be noted here that the status statistics such as percentiles and means were computed using the beginning of the semester sample size while the parts of the study which determined the relationship of reading to general academic achievement utilized the net end of the semester sample of 186.

The Analyses

The preliminary analysis consisted of computing the basic descriptive statistics which consisted of the means and standard deviations for reading test scores, success point averages, and age. Analyses subsequent to these descriptive statistics corresponded to phases I, II, and III of the study.

Phase I undertook to determine if there was an overall general relationship between reading test scores and academic achievement. This phase was a necessary first step to determine if it were warranted to continue the study;

this phase essentially served to determine whether or not the NDRT was a relatively valid measuring device - the major purpose of Phase I.

In order to determine this general relationship between test scores and general academic achievement, the vocabulary (V), comprehensive (C), and the total (T) scales of the NDRT were each correlated with the criterion, which was the success point average (spa). The Pearson produce-moment coefficient of correlation statistic was used for this purpose. It was also decided to determine the relationship between age and general academic achievement using the same statistic to determine how much maturity affected academic achievement; therefore, age was also correlated with the criterion - spa.

Phase II was conducted to establish a cutoff score for reading placement. This score was to be the one at or below which students making that score would be required to enroll in a reading improvement course, and scoring above which they would not be so required. The reading test scale of the NDRT utilized for this purpose was the total (T) scale, which is the combined vocabulary and comprehension scores (V + C). The T scale had been found to be more related to the criterion than either the V or the C scale alone.

Twenty-five T scores were tried out as hypothetical cutoff scores in order to determine the most efficient one. These scores ranged from slightly more than one standard deviation below the mean to approximately a half a standard deviation above the mean.

Using a 2 x 2 contingency table, a phi coefficient (ϕ) and a chi-squared statistic (x^2) were calculated for each of the 25 trial cutoff T scores. The phi coefficients provided a relative index of relationship between the trial cutoff scores and academic success. Success was defined here as the achievement of a 2.000 spa, on a four point grading scale. A "success" was any spa of 2.000 or greater. Any success point average below a 2.000 was tallied as a "non-success" in the 2 x 2 table.

Because the phi coefficient underestimates relationship, each phi was converted to a tetrachoric coefficient of correlation (r_{tet}) as demonstrated by Wert, Neidt, and Ahmann (1954). The tetrachoric coefficient approximates the more familiar Pearson r (Guilford, p. 305), and may be interpreted in a manner sililar to the more familiar r. The chi-squared statistic was used here as a test for statistical significance. The procedure described here was essentially similar to that used by the U. S. Employment Service to develop occupational aptitude patterns for the General Aptitude Test Battery (1958).

Phase III, the feedback phase, was included for utilitarian purposes to provide more meaningful feedback than heretofore to the students tested by the NDRT and to other users (faculty, administrators, and others) as well. In addition to simply being informed as to whether or not one would be required to enroll in remedial reading, which itself is a type of feedback, it was decided that it would be very helpful to students to know how one stands in relation to ones peers; thus, in Phase III were developed local norms for the total (T) scores in the form of percentile ranks. Also, for each total NDRT raw score a predicted semester success point average was calculated. In all, the total feedback to the student could include whether or not the student had to take remedial reading, where he/she ranked in relation to his/her peers on the national level (from the NDRT Manual), where she/he ranked on the local level (two year public community college), and what his/her expected first semester grade point average might be, corrected to account for overall "success" by taking withdrawals into account (success point averages - spa's).

The predicted success point averages were computed using the linear regression equation $y=b+mx$ in which y is the functionally dependent variable (success point average) and x is the independent variable (the raw T score on the NDRT). The total equation may be interpreted as a straight line fitted to the sample points (x, y). The line would have b as its y intercept and a slope of m.

In order to provide for convenience of feedback to students, the feedback information was placed into a table. Included are the raw T scores of the NDRT and a corresponding national percentile, a local percentile, and the predicted success point average for each T score. Two tables were constructed, one for both forms of the NDRT to display the feedback information. The second table, for Form D scores, was interpolated from form C.

RESULTS

Table 1 indicates that for the sample used here the mean vocabulary (V) score of the Nelson Denny Reading Test (NDRT) was 24.9, which according to the "national" norms for the 13th grade, is the 31st percentile; the means for the comprehension (C) and total (T) scales were in the 32nd and 28th percentiles respectively according to the "national" norms. Thus, the "average" community college student in this sampling falls somewhere between the lowest third and the lowest quarter nationally. Two aspects of this finding should be noted. First, this finding is not atypical when comparing open admissions public two year college student populations with samples that are "national" in scope, containing as they do, sizable components from selective colleges and universities. Secondly, it should be reiterated that the so-called "national" sample as reported in this NDRT Manual (1973) was not a

stratified random sampling of all the various types of colleges; the sample used to obtain their national sample was obtained according to whatever institutions could be found to cooperate, and the numbers of students from the various types of institutions were not selected so as to be proportionally representative of a true national sample of college freshmen.

Table 1

Descriptive Statistics

N=186

Variable	Mean	SD	Grade 13 National Percentile
Vocabulary (V)	24.9	11.429	31
Comprehension (C)	34.5	10.124	32
Total (T) (V + C)	59.3	19.348	28
Success Point Average (spa)	2.130	.942	NA
Age	19	2.883	NA

Table 1 shows that the mean T scale score was 59; Figure 4 indicates the median was 57. Thus, the median and the mean were nearly the same, indicating that the distribution was nearly normal, with a slight positive skew. Figure 3 represents this graphically. Figures 1 and 2 show similar results for the V and C scale scores. Figure 3 also states that the national median raw score for the T scale was 73. Thus the "national" median was the 80th FV percentile; the FV median of 57 was between the 25th and 26th percentile nationally. Again, that was not unexpected, nor particularly meaningful, given also the relative non-national quality of the published "national" norms.

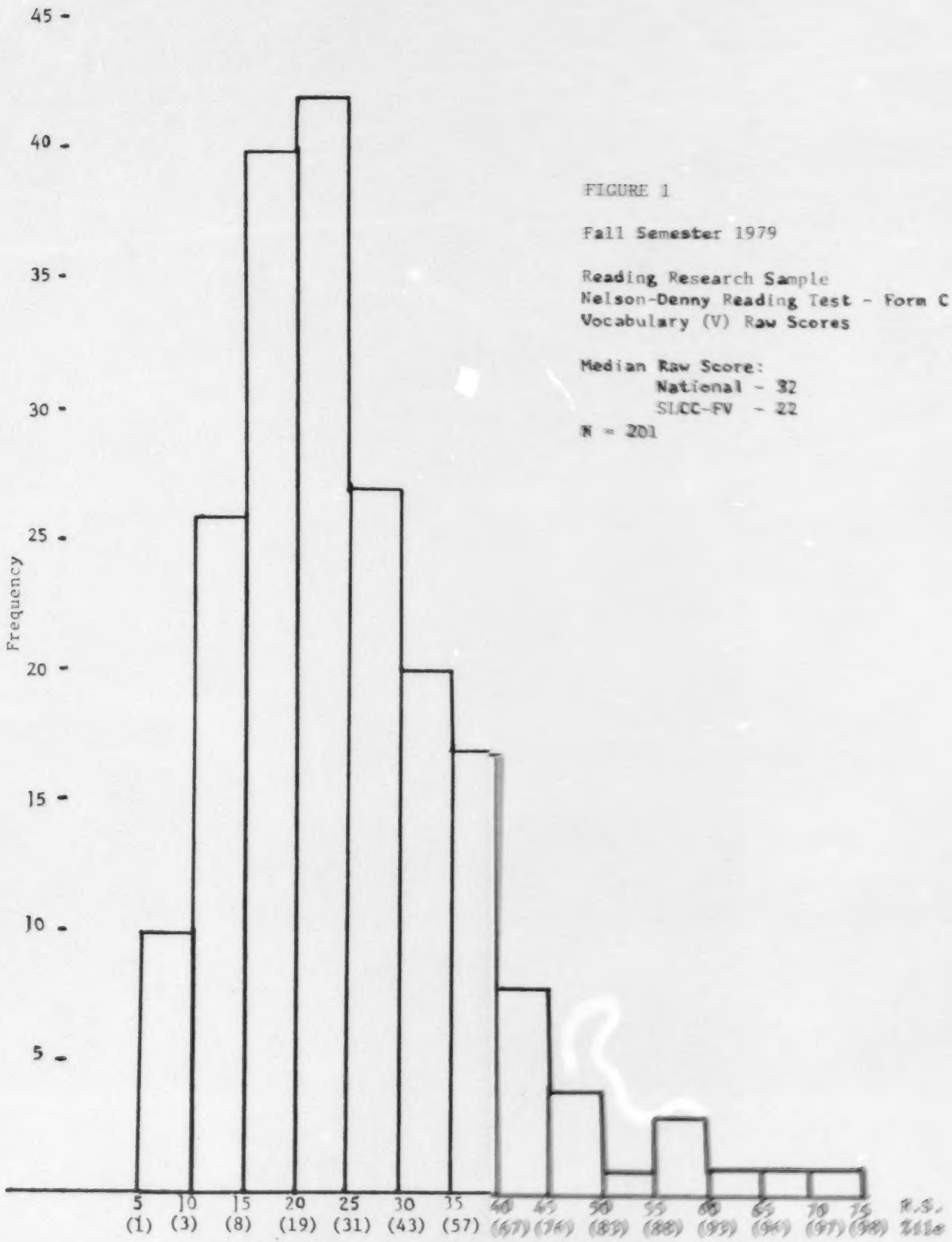

FIGURE 1

Fall Semester 1979

Reading Research Sample
Nelson-Denny Reading Test - Form C
Vocabulary (V) Raw Scores

Median Raw Score:
National - 32
SLCC-FV - 22
N = 201

30

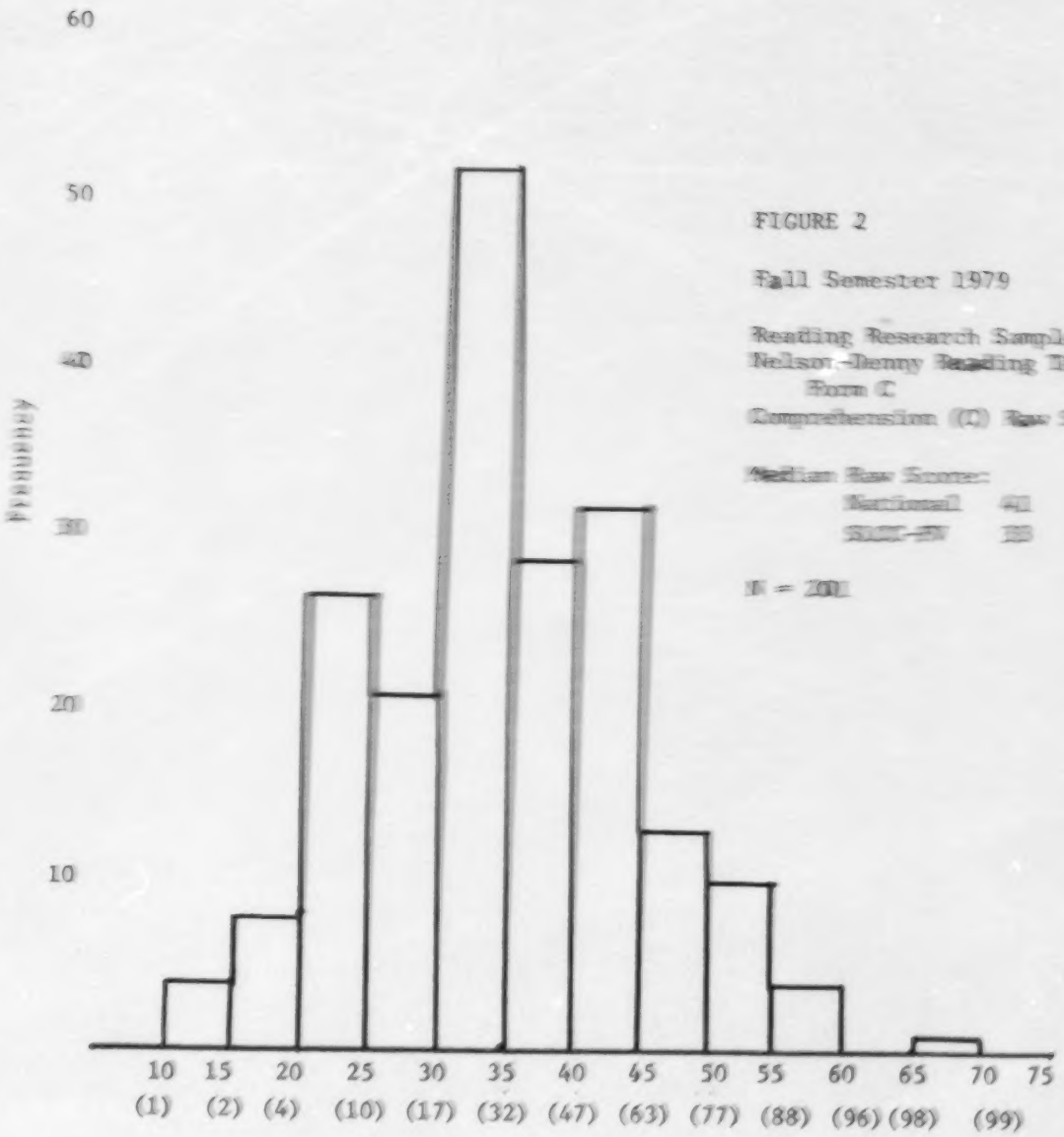

FIGURE 2

Fall Semester 1979

Reading Research Sample
Nelson-Denny Reading Test
Form C
Comprehension (C) Raw Scores

Median Raw Score:
National 41
[illegible] [illegible]

N = 200

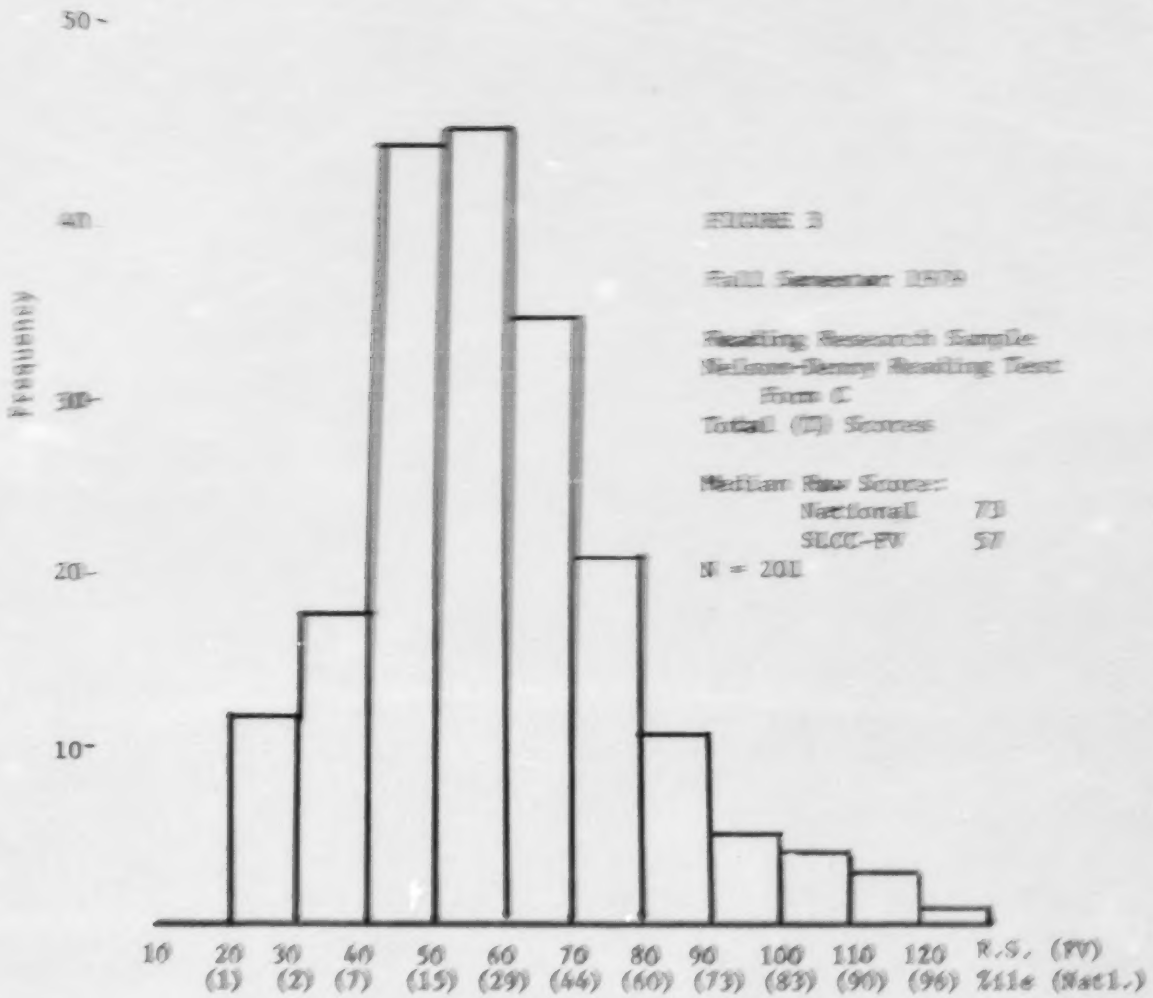

FIGURE 3

Fall Semester 1979

Reading Research Sample
Nelson-Denny Reading Test
Form C
Total (T) Scores

Median Raw Scores:
National 73
SLCC-FV 57
N = 201

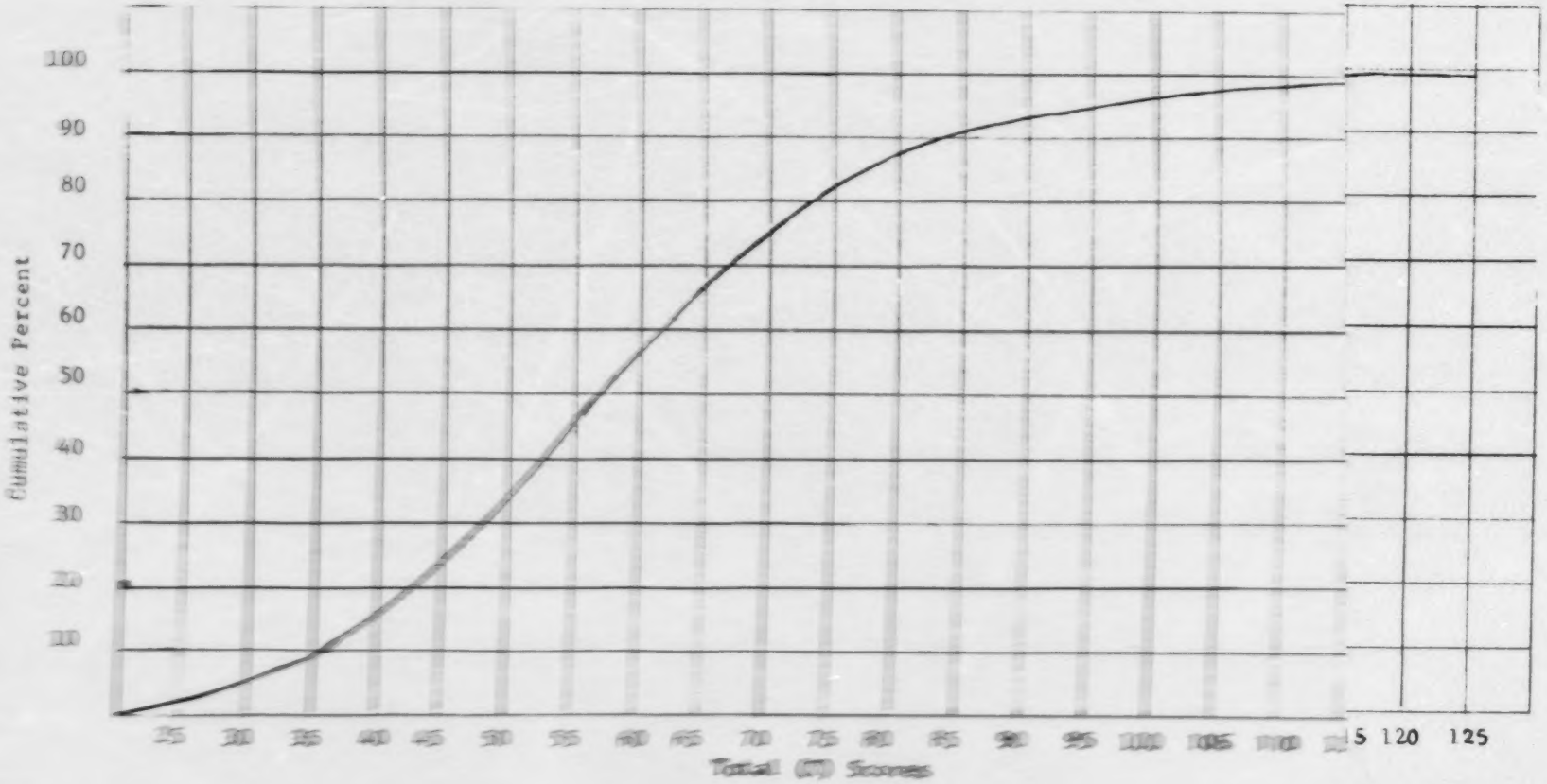

FIGURE 4
Nelson-Denny Reading Test

Only means and standard deviations are reported in table 1 for success point averages and age. No national age scores were reported in the manual and obviously no national success point averages were available either.

It was interesting to note that the mean grade adjusted for withdrawals was 2.130, an almost perfect "C," or average, grade. This finding tends to further confirm the procedure of making the adjustments for withdrawals. It is also very interesting that the mean spa for students taking multiple courses (this study) was almost identical to the mean grade found in another similar study which used the same sample, but selected only one course per student instead of an overall grade average as the criterion (Hartman, Bagwell, 1980). The mean for one single course, on a four point scale, was 2.108; for all courses enrolled in and completed, the mean was 2.130. The standard deviations were 1.193 (single course) and 0.942 (multiple course). In the single course study, the single course for each student in the sample was selected so as to be one in which reading was the primary academic skill required, as opposed to other skills such as mathematics, writing, art, music, or whatever. The major reason for doing a follow-up analysis was that it was believed that such a study would more nearly approximate "a real life situation" for first-time college students and thus would be more generalizable. It would also be more accurate in that the criterion represented a much larger sampling of grades assigned; that is, the single grade criterion sample was 201, the same as the number of students. The multiple grade (spa) sample probably contained three or four times that number. That would also probably explain the smaller variance for the multiple course study. Extreme grades would tend to cancel each other out.

The following paradigm is a synoptic comparison of the mean grades and S.D.'s for the single course study and the multiple course study.

Criterion	Mean	S.D.
Single Course (One grade/student)	2.108	1.193
Multiple Course (Success Point Average)	2.130	0.942

Here it is seen that the success point average mean was slightly higher than the mean for single courses. That may be somewhat surprising, in that W's were converted to F's in the spa study; however, the single courses in the first study were selected so as to represent fairly rigorous academic courses, as opposed to developmental courses, as other skills courses. This result also generally highlights again the notion that the present grading system which allows totally for non-punative withdrawals will result in somewhat more inflated grade point averages than would be the case at colleges with a more traditional withdrawal policy.

At any rate, the observation that the mean spa was a "C" average further confirmed the procedure of using the spa rather than the gpa.

Table 1 indicates that the V and C scores were very nearly the same in that the mean raw score for V was in the 31st percentile and the mean raw score for C was in the 32nd percentile, according to national norms. The NDRT manual interprets closeness of V and C scores to mean that students do not have unusual or motivational problems. Marked differences between V and C are reported to be indicative of motivational type problems.

Finally table 1 indicates that with respect to age the group sampled here represents what may be thought as the "typical college freshman." The

mean age was 19, with a standard deviation of three, indicating that most of the students sampled were in their late teens to early twenties. This finding is no doubt because all the students were daytime students, and most were also attending full time.

Phase I - The Validation Phase

Table 2 provides a succinct overview of the relative degree of relationship between reading skill and academic achievement, as was previously defined. Also, shown in parentheses are the same relationship indicators using only the single course/reading skill relationship (Hartman, Bagwell, 1980).

Table 2

Coefficients of Correlation Between

Reading Test Scores and Semester spa

N=185

	Coefficient of Correlation r	Coefficient of Determination r^2	Sig. Level
Vocabulary	.42 (.37)*	.18 (.13)	.01
Comprehension	.40 (.42)	.16 (.18)	.01
Total	.55 (.44)	.30 (.19)	.01
Age	.06 (.09)	.004 (.01)	Not Sig.

* The numbers in parenthe[illegible] are from the single course criterion study (Hartman & Bagwell, 1980).

The most important finding here is probably that the coefficient of determination (r^2) for the total (T) score scale was considerably higher for the multiple course criterion study than for the single course study - .30 compared to .19. In general terms, then, reading skill in the present study

accounted for about 11 percent more of the variance in academic achievement than reading skill (as measured by the NDRT) did in the former study which only used one course per student as the criterion. Or, this may roughly be interpreted to mean that reading skill level as measured by the NDRT accounts for almost one-third of the academic achievement variance in the "real life" multiple course study as opposed to only about one-fifth in the more artificial single course criterion study. It is seen that the correlations for NDRT and achievement were highly significant statistically. It was also reassuring that age, in this study, was not a significant factor in academic achievement, showing that only .4 of one percent of the variance in achievement could be attributed to variability in age, and that indicator of relationship (age) could easily have been due to chance sampling error.

The correlation of .55 between the NDRT T scale score and scholastic success found here was better than most of those reported in the NDRT manual. Of 94 reported in the manual (p.14), the median correlation was only .47; the range was .16 to .67. Raygor (1978) stated:

> Some predictive validity data now reported suggest that scores are somewhat predictive of college grades, even though some of the generalizations made are questionable. For example, the manual indicates that in one sample the NDRT was "slightly better" than the Minnesota Scholastic Aptitude Test in prediction - .397 and .384 respectively, and both were lower than high school rank at .585 (p.1209, emphasis added).

It is reassuring to note that the .55 in the present study was considerably higher than either of the relationships reported by Raygor for the NDRT or another test, and was nearly as high as high school rank at the University of Minnesota. High school rank in recent years has been found to be relatively unreliable and invalid as a predictor of scholastic achievement for students in open admissions institutions.

The results from the validity phase (Phase I) were at least as good, if not better than were expected. Thus, it was felt that further work to complete phases II and III was warranted.

Phase II - Reading Requirement Cutoff

Phase II was the focal point of this study and was initially the only reason the study was undertaken.

Table 3 on the following page provides comprehensive data which shows what would have happened (to students) if any of 25 different NDRT T scale scores had been used to determine who would have been required to go into a reading improvement program. Each score in the first column represents a hypothetical cutoff score, at, or below which, a student making that score would be required to enroll in four semester hours of non-transferable course work in reading improvement. For some students that may be an odious requirement to be avoided adamantly at all cost.

Because the NDRT scale accounted for more of the variance than either the V or C scales, the T scale was selected for trial cutoffs. In analyzing the various trial cutting scores, a cutoff of "pass" or "not pass" had to be determined. A 2.000 spa was considered the minimum passing grade average for the criterion.

In order to facilitate the use of Table 3, each of the 14 columns is explained as follows, beginning with column 2. The raw score of 35 is used throughout the example to illustrate Table 3.

Column 2, "Nat. %." The national percentile equivalent of the given raw score; 35 is the fourth percentile.

Column 3, "Loc. %." This is the local (Florissant Valley) percentile. In this example, the 9th was the percentile, locally, for a score of 35.

Table 3

NDRT Cutoff Analysis
with a 2.000 Success Criterion
Form C Nelson-Denny

Cutoff			Hits			MISSES							
	Nat	Loc			Total			Total	Total Screened	% Pass who were screen			
T	%	%	Above Pass	Below NP	# / %	Below Pass	Above NP	# / %	# / %		χ^{2*}	β	r_{tet}
35	4	9	106	14	120 / 65	3	63	66 / 35	17 / 09	18	12.936	.264	.40
38	6	13	103	16	119 / 64	6	61	67 / 36	22 / 12	27	10.095	.233	.36
40	7	16	103	16	119 / 64	6	61	67 / 36	22 / 12	27	10.095	.233	.36
42	8	19	103	21	124 / 67	6	56	62 / 33	27 / 15	22	17.231	.304	.46
43	9	21	102	22	124 / 67	7	55	62 / 33	29 / 16	24	16.821	.301	.46
44	10	22	101	27	128 / 69	8	50	58 / 31	35 / 19	23	22.706	.349	.52
45	11	25	99	28	127 / 68	10	49	59 / 32	38 / 20	26	20.520	.332	.50
46	12	27	96	34	130 / 70	13	43	56 / 30	47 / 25	28	24.821	.365	.54
47	13	29	95	36	131 / 70	14	41	55 / 30	50 / 27	28	26.397	.377	.56
48	13	31	94	39	133 / 72	15	38	53 / 28	54 / 29	28	29.801	.400	.59
49	14	34	92	42	134 / 72	17	35	52 / 28	59 / 32	29	31.606	.412	.60
50	16	36	89	43	132 / 71	20	34	54 / 29	63 / 34	32	28.323	.390	.58
51	17	38	88	44	132 / 71	21	33	54 / 29	65 / 35	32	28.476	.391	.58
52	18	40	84	46	130 / 70	25	31	56 / 30	71 / 38	35	25.899	.373	.55
53	20	43	81	48	129 / 69	28	29	57 / 31	76 / 41	37	25.082	.367	.55
54	21	46	80	48	128 / 69	29	29	58 / 31	77 / 41	38	23.748	.357	.53
55	23	47	78	51	129 / 69	31	26	57 / 31	82 / 44	38	26.147	.375	.56
56	24	49	77	52	129 / 69	32	25	57 / 31	84 / 45	38	26.552	.378	.56
57	26	51	77	53	130 / 70	32	24	56 / 30	85 / 46	38	28.233	.390	.58
58	27	53	73	55	128 / 69	36	22	58 / 31	91 / 49	40	26.629	.378	.56
59	28	55	66	55	121 / 65	43	22	65 / 35	98 / 53	44	18.512	.316	.48
60	29	58	62	57	119 / 64	47	20	67 / 36	104 / 56	45	17.486	.307	.46
62	32	62	58	60	118 / 63	51	17	68 / 37	111 / 60	46	18.176	.312	.47
64	35	65	55	64	119 / 64	54	13	67 / 36	118 / 63	46	21.932	.343	.51
66	39	69	48	67	115 / 62	61	10	71 / 38	128 / 69	48	20.273	.330	.50
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.

*Probability >.99 for All χ^2s.

Column 4, "Above Pass" - 106 students scored above the trial cutoff of 35 and obtained a 2.000 or higher, spa - were categorized "pass."

Column 5, "Below NP." - 14 students scored below 35 and did not pass - obtain an spa of at least 2.000.

Column 6, "Total #/%." Total number of "Hits." All students who scored above the cutoff and achieved a 2.000 spa, plus all who scored at or below the cutoff and did not achieve a 2.000. For the score of 35 there were 120 hits which means that for a raw score of 35, 65% of the sample would have been screened correctly - hits.

Columns 7, 8 and 9 are the other side of the coin - the misses.

Column 7, "Below Pass." 3 students scored below the hypothetical 35 cutoff, would not have been allowed to skip developmental reading, but would have achieved a 2.000 spa.

Column 8, "Above NP." 63 students scored above 35, but even so did not achieve a 2.000 spa. (This would possibly indicate the cutoff should be higher.

Column 9, "Total #/%." This represents the sum of columns 7 and 8, or total number and percent of misses. For a raw score of 35, 66 students, or 35 percent of the sample would have been misses.

Column 10, "Total Screened - #/%." The total number and percent of sample who scored below the cutoffs - regardless of whether they were "Hits" or "Misses."

Column 11, "% Pass who were screened." This is column 7 (Below pass) divided by column 10, the number schreened by a cutoff. For a

cutoff of 35, 17 students scored at or below 35; of those 17, 3 achieved a 2.000 spa (passed). Thus, 3/17 or 18% of those screened by a score of 35 would have been "penalized." It is seen that as the cutoff is raised the percent penalized increases.

Column 12, "X^2." This is the chi squared statistic used to determine probability. It may be noted that all chi squares were highly significant statistically.

Column 13, "Ø." This is the phi coefficient which was derived from chi squared. In its present form it shows only a relative degree of relatedness between the cutoff score and efficiency of screening - the score which has the most hits and the least number of misses.

Column 14, "r_{tet}." This is the telrachoric correlation which is also, like phi (Ø), an indicator of relatedness of between the magnitude of the cutoff score and number of hits/misses. In addition, however, r_{tet} may be interpreted similarly to the better known Pearson product - moment coefficient of correlation.

From Table 3 the most efficient T score cutoff appears to be 49. Using it there would be the highest ratio of hits to misses. The r_{tet} of .60 is similar to the r of .55 between the NDRT raw scores and general scholastic achievement. The 49 raw score is the 14th percentile on "national" norms and the 34th percentile on Florissant Valley norms; that means, then, that approximately one-third of full-time day students could realistically be required to enroll in a reading improvement program. (Column 10 shows that the percent who scored at or below 49 of the sample was 32%; this apparent

discrepancy - between 34% and 32% - is because the figures in Table 3 exclude those 15 students of 201 who withdrew totally. Thus, Table 3 is based on an N of 186, whereas the FV percentile ranks are based on 201.)

Although a cutoff score of 49 yielded an r_{tet} of .60 and the ratio of hits (134) to misses (52) for that score is 2.6:1, the negative factors inherent in this, the best, cutoff should also be noted. Nearly one-third, 29%, of the students who would have been required to take developmental reading could have achieved a 2.000 spa in a full academic load without having taken the reading program. (Had a gpa been used instead of the spa, the percentage of successes who scored at or below 49 would have been even higher.)

Table 4 depicts rather graphically what happens as screening cutoff scores are raised and lowered.

As a score is raised, obviously so do the academic "successes" of those who score above it. So, why not set the score quite high and be better able to "guarantee" success?

Because, as the cutoff goes up, and the success rate does too, so also does the percent rate of people who did not achieve the cutoff but did achieve the academic success. That is, as the cutoff is raised, the "penalty factor" is increased.

Here it is clear that as the cutoff is raised the percentage who "passed," (got a 2.000 spa or better) increased: The percentages of "success" goes from 63% to 83%, as shown in column 2. Likewise, as the cutoff goes from 35, to 49, to 66, the percent of those who score above the cutoff and who "fail" decreases: 37, 28, and 17 percent respectively (col. 3).

Table 4
Cost - Effectiveness Rates
for the
Lowest, Highest, and Optimum Cutoffs

	Raw Score	% Above who passed	% Above who "Failed"	% Below who Passed
Lowest Cutoff	35	63	37	18
Optimum Cutoff	49	74	28	29
Highest Cutoff	66	83	17	48
Column No.	(1)	(2)	(3)	(4)

Thus from columns two and three above it might be concluded that the higher the cutoff score is set, the better: more successes and fewer failures. But, column 4 shows that there is a cost: only 18% who scored below the cutoff would have passed anyway when the cutoff was 35; 29% below the cutoff of 49 (optimal cutoff) would have been successful in spite of being below the cutoff; and almost half, 48%, would have achieved at least a 2.0 gpa even though they were below the cutoff, had the cutoff been set at 66 - the highest trial cutoff. Thus, Table 4 should cause decision makers to weigh carefully the human cost factor before establishing any score as a cutoff that causes people to do whatever it is that is required - mandatory - which some people might not otherwise freely do.

Phase III - The Feedback Phase

Whenever institutions of education require students be tested en mass, good test procedure dictates that those tested be given as complete feedback as possible, including adequate opportunities for counseling and interpretation of the test scores. Merely to inform the student that "you are required to do thus and such because you scored at or below a certain score on a test,"

is very inadequate feedback. It is also inadequate for those students who are not required to do whatever if they are told simply that they are not required to do whatever because they "passed the test."

Students want to know and need to know how they did on the test in relation to their peers; they also need, and often want to know, what their scores predict about their future performance - on whatever it is the test is presumed to relate to.

Thus, in order to provide as much such feedback as possible, an appendix has been attached which provides "national" percentiles as well as local Florissant Valley percentiles for the total raw (T) scores for both the D and the C forms. In addition, for each raw T score is a predicted success point average (spa), which will always be as high, or higher than the institution's gpa.

Armed with such information, a counselor should be able to provide the students comprehensive feedback which could be used as a basis for more meaningful counseling and advising. Of course, knowledge of whether or not the student scored above the cutoff requirement is also a form of feedback and would be a part of the total feedback provided the student.

SUMMARY and DISCUSSION

This study was an outgrowth of the establishment of a reading improvement program that is mandatory for all regular full-time students entering college for the first time and whose reading skills were judged to be so minimal that they were given little or no probability of achieving academic success.

In order to implement the reading requirement, it was necessary that the college administer a reading test to students whose general placement tests

evidenced poor reading skills. It was then necessary to set a cutoff score on the reading test which was a score, at or below which the students would be required to enter into the program of reading improvement - four credit hours of non-transferable course work. The problem, to be addressed by this study, was what score on the test should be used as the cutoff. Implicit in this problem was the one of test validity. Many test users felt the test itself was faulted, and to use it with whatever cutoff resulted in injustices to many students. Thus, before addressing the problem of recommending a specific cutoff score on the test, the test itself needed to be validated, using a sampling of the population which was being tested - an open admissions community college. Finally, if the test were found to be sufficiently valid so as to warrant its continued use, and if the most efficient (fair) cutoff could be established, it was decided that users needed more and better information about the meaning of test scores. That is, more and better feedback data were needed to improve the use of the test for all who took it by providing more knowledge than merely informing the student whether or not reading improvement would be required.

Initially some 500 students were tested during the first week of a semester. After eliminating approximately 300 students in an attempt to control for numerous variables, 201 students were included in the "final" beginning of the semester sample. The normative data, such as percentiles, means, and the like, were based on the 201 figure. During the course of the semester, 15 students withdrew completely; therefore, the final criterion data and the cutoff and prediction statistics sample number was 186.

Judging from a survey of the literature regarding The Nelson Denny Reading Test (NDRT), and based on local experience, it was somewhat surprising to find that the NDRT correlated with academic achievement as well as it did.

The Pearson r was .55 between the total (T) raw score scale and spa-success point average, which is a modified gpa. The gpa modification consisted of assigning zero grade points to course withdrawals, whereas the institution does not average W's.

Of the studies reported in the NDRT Manual (1973), the median correlation between the NDRT and scholastic achievement was .47; thus, the .55 found here was well above the median of those reported in the Manual. It was all the more surprising to note that the Manual reported a .58 correlation between high school class rank and scholastic success at the University of Minnesota, in an era when class rank was the best single predictor of college success. Thus, the .55 found here is almost as good a predictor as class rank was back when it was a good (the best) predictor.

Armed with finding that the NDRT was reasonably valid relative to the population concerned, it seemed well worth recommending the continued use of the NDRT in general, and well worth proceeding to find a useful, accurate, and fair cutoff score on the test.

The most efficient cutoff score, of 25 different ones tested, was a total (T) score of 49, Form C. (The percentiles for form D indicated that a raw score of 49 on both forms was the same.) This score, 49, was five raw score points higher than the 44 being used by the reading department and which had been found to be the most efficient cutoff in a previous related study in which only a single course grade per student was used as the cut-off (Hartman and Bagwell, 1980). According to "national" norms published by the NDRT manual, a raw score total of 44 is the 10th percentile; a score of 49 is the 14th. Using the local norms, found in this study, the previously determined cutoff of 44 is the 23rd percentile; the new one (49), based on a total course load per student is the 34th percentile. That is this study

suggests that one-third of the entering students do not have a realistic chance to obtain the minimally acceptable 2.000 spa without improving their reading skills. Thus, by Cohen's (1976) definition, those students may be thought of as functionally illiterate.

On the basis of this study, which used what is probably a more realistic criterion than just the one course, it is recommended that the cutoff be revised upward to the raw score of 49, which would mean that approximately 11% more (34%) of the incoming students would "qualify" for reading improvement. This may not be as discouraging as it first appears when it is remembered that Rouche (1978) claimed the average community college student is functionally illiterate - again, by Cohen's (1976) definition. By this study, for this population, "only" one-third might be so termed.

The final phase of this study consisted of generating feedback data. These are located in tabular form in the appendix. For each total raw score on both form C and D are the following: The "national" percentile, the local (FV) percentile, and a predicted gpa - modified, as previously described, to reflect a more accurate index of scholastic achievement than the gpa currently in use by the institution.

In evaluating this study, and in generalizing the results, there are several important factors that should be noted carefully. These are mentioned below.

A word of caution is in order regarding the age of the subjects of this study relative to the age of the general student population of Florissant Valley. The mean age in the study was 19 years; the median age of the college is in the 27-28 year range. Although the median and the mean may not be the same, caution should be used in generalizing these findings to the more mature

evening student who probably also has a full-time job, a family, who attends classes part-time, and who, by and large, is more highly motivated academically.

Another area of possible interest that was brought into relief by this study was the course withdrawal question. It was found, first of all, that students who withdraw totally from all courses are not significantly different in reading skill level from those who remain in all, or most of their course-work. This finding should not be forgotten in any subsequent policy change regarding withdrawals. Thus, factors other than reading skill, for the most part, appear to be the reason for withdrawal from all courses by a student. Secondly, it was discovered that the "average" student receives the "average" (C) grade when individual course withdrawals are treated as F's. That is, the mean gpa when W's were counted into the gpa as contributing zero (0) grade points was 2.1 - about as close to a 2.000 (on a four point scale) as would be expected by chance sampling error; in fact, it was just slightly higher than a perfect 2.000. Thus, by not allowing any penalty, at any point in a semester for a W, the gpa's for students who withdraw to avoid lowered gpa's is obviously higher than those of students who persevered; thus, under the present withdrawal policy, the meaning of the gpa itself becomes distorted, less meaningful, and for the institution as a whole, supiously high. Many receiving institutions and employers are probably aware of this policy; this policy, therefore could negatively affect the overall image of the institution - its credits and its graduates. If so, the policy might bear re-examination by the policy making powers that be.

Another point this study graphically illustrated, and which should be emphasized is that whenever a test score is to be used as a screening device to require students to do whatever, there will be a penalty factor at work because (A) no test is perfect and (B) no criterion (that which is being

predicted) is absolutely stable in human institutions, as would be the case in the physical sciences: for exapmple, there is an absolute relationship between the temperature and the volume of gas. No such absolute relationship exists between human reading skill as measured by any humanly devised test and human learning (academic) achievement measured by any grade or grade point average assigned by humans to other humans. Also, in educational research, the number scales are assumed to be ratio scales, when in fact, unlike the physical sciences, they are far from it. Therefore, because perfect prediction is impossible, perfect placement is impossible.

Tables 3 and 4 (results section) depict what occurs as a cutoff score is raised or lowered in terms of "penalties" to students. These tables should be carefully scrutinized and the meaning of their contents recognized by all persons (including committees) involved in setting requirements, or prohibitions, applied to other humans on the basis of test scores - whether they be reading test scores for reading placement, placement test scores for college entrance, or whatever else. It should be apparent from these tables that to raise a cutoff higher than absolutely necessary (the one that is the most efficient) subjects an unnecessary number of persons to the liklihood of being penalized (unfairly discriminated against). To reiterate, even the most efficient cutoff entails a certain waste of human potential and the risk of lowering self-esteem attendant to that waste. It is already hotly debated in some quarters whether or not even the most efficient cutoffs are ethical - especially as applied in that "handmaiden of American democracy," the public, tax supported community junior college. To avoid further criticism, therefore, cutoffs no higher than are indicated by empirical research would seem to be justifiable

It is greatly hoped that all the faculty, administrators, board members, or whoever else sets or influences testing policy and procedures, and that whoever uses tests with students will find this study informative and helpful, and that ultimately the main benefits will accrue to the students.

REFERENCES

Abel, J. Academic advising: goals and a delivery system. Journal of College Student Personnel, 1980, 2, 151-153.

Buros, O. K. The eighth mental, measurements yearbook. Highland Park, N. J.: The Gryphon Press, 1978.

Cohen, A.M. Shall we segregate the functionally illiterate? Community and Junior College Journal, 1979, 15, 14-18.

Cross, K. P. Accent on learning. San Francisco: Jossey-Bass, 1976.

Davis, F. In O. K. Buros (Ed.) The eighth mental measurements yearbook. Highland Park, N. J.; The Gryphon Press, 1978.

Forsyth, R. A. In O. K. Buros (Ed.) The eighth mental measurements yearbook. Highland Park, N. J.: The Gryphon Press, 1978.

General Aptitude test battery, guide to the use of the: Section III - development. Washington, D. C.: Bureau of Employment Security, 1958.

Guilford, J. P. Fundamental statistics in psychology and education. New York: McGraw-Hill Co., Inc., 1956.

Hakstian, R. A. In O. K. Buros (Ed.) The eighth mental measurements yearbook. Highland Park, N. J.: The Gryphon Press, 1978.

Hartman, N. E. and Bagwell, M. The empirical determination of a minimal test score necessary for probable success in a transfer level course for which reading is the primary academic skill. St. Louis: St. Louis Community College at Florissant Valley (unpublished report), 1980.

Nairn,A. and Associates. The reign of ETS: The corporation that makes up minds. Washington, D. C.: 1980.

Nelson-Denny reading test, the examiner's manual. J. I. Brown, M. J. Nelson, and E. C. Denny. Boston: Houghton Mifflin Co., 1973.

O'Connor, M. "City colleges plot revolution - reading standards." Chicago: Chicago Tribune, Section I, p. 3, 1980.

Raygor, A. L. In O. K. Buros (Ed.) The eighth mental measurements yearbook. Highland Park, N. J.: The Gryphon Press, 1978.

Student body profile, fall 1980. Unpublished report, Office of Development and Research, St. Louis Community College, St. Louis, Mo., 1980.

Trillin, A. S. and associates. Teaching basic skills in college. San Francisco: Jossey-Bass, 1980.

Wert, J. E., Neidt, C. O., and Ahmann, J. S. Statistical methods in educational and psychological research. New York: Appleton-Century-Crofts, Inc., 1954.

APENDIX

NATIONAL AND FLORISSANT VALLEY NORMS

FOR

THE NELSON-DENNY READING TEST

For Total (V + C) Raw Score Only

A. Column 1.

Total (V + C): Find total raw score (V + C) in the first norm column.

B. Column 2.

Natl. %ile: This is percentile rank based on a National 13th grade sampling for first-time college students. This was derived from a mix of two-year and four-year college students; from the Nelson-Denny manual.

C. Column 3.

FV %ile: Percentile rank for SLCC-FV students. N = 185, first time (13th grade) students. Fall semester 1979.

D. Column 4.

FV gpa_n: This is a "net" predicted semester grade point average (gpa_n) for the first semester at SLCC-FV for full-time students, or students for whom job plus work equals full-time.

Note: The gpa_n was determined by recalculating the semester gpa, and assigning zero grade points to W grades.

These predictions were derived using the linear regression equation; the correlation coefficient was .55 - significant beyond the .01 level.

NH/djm
5-22-80

NELSON-DENNY READING TEST NORMS
Grade 13, Beginning of Year - Form D*
Prepared by
Neal E. Hartman
ST. LOUIS COMMUNITY COLLEGE AT FLORISSANT VALLEY

Total (V+C)	Natl. %ile	FV %ile	FV gpa_n	Total (V+C)	Natl. %ile	FV %ile	FV gpa_n	Total (V+C)	Natl %ile	FV %ile	FV gpa_n	Tota (V+C)	Natl. %ile	FV %ile	FV gpa_n
51	16	36	1.880	80	59-60	87	2.655	109	80	98	3.511				
50	15	35	1.880	79	57-58	86	2.629	108	90	98	3.457	137			
49	14	32,33 34	1.854	78	56	85	2.602	107	89	98	3.431	136	99	99	4.000
48	13	30-31	1.827	77	54-55	83-84	2.575	106	89	98	3.404	135	99	99	4.000
47	13	28-29	1.800	76	52-53	82	2.548	105	88	97	3.377	134	99	99	4.000
46	12	26-27	1.773	75	50-51	81	2.503	104	87	97	3.350	133	99	99	4.000
45	11	24-25	1.747	74	48-49	79-80	2.468	103	86	97	3.335	132	99	99	4.000
44	10	22-23	1.720	73	46-47	77-78	2.442	102	86	97	3.324	131	98	99	4.000
43	9	20-21	1.693	72	44-45	75-76	2.415	101	85	97	3.297	130	98	99	4.000
42	8	19	1.666	71	42-43	72-74	2.388	100	84	97	3.270	129	98	99	3.996
41	7	17-18	1.640	70	40-41	70-71	2.361	99	83	97	3.232	128	98	99	3.991
40	7	16	1.613	69	39	69	2.335	98	82	96	3.190	127	97	99	3.938
39	6	14-15	1.586	68	37-38	67-68	2.308	97	80-81	96	3.150	126	97	99	3.912
38	6	13	1.560	67	35-36	64-66	2.281	96	79	96	3.110	125	97	99	3.885
37	5	11-12	1.533	66	34	63	2.254	95	78	95	3.083	124	97	99	3.858
36	5	10	1.506	65	33	63	2.228	94	77	95	3.056	123	97	99	3.831
35	4	9	1.485	64	32	61-62	2.201	93	76	95	3.030	122	96	99	3.817
34	4	9	1.479	63	31	60	-	92	74-75	94	2.985	121	96	99	3.805
33	4	8	1.453	62	30	59	2.174	91	72-73	93-94	2.949	120	96	99	3.778
32	3	7	1.426	61	29	56-58	2.148	90	71	92	2.896	119	96	99	3.751
31	3	7	1.372	60	28	54-55	2.121	89	70	92	2.869	118	95	99	3.725
30	2	5	1.346	59	26-27	51-53	2.094	88	69	92	2.843	117	95	99	3.698
29	2	4	1.319	58	25	50	2.067	87	68	91	-	116	94	99	3.686
28	2	4	1.292	57	24	48-49	2.041	86	67	91	2.816	115	94	99	3.671
27	1	3	1.266	56	22-23	46-47	2.014	85	66	90	2.789	114	93	99	3.644
26	1	3	1.239	55	21	44-45	1.987	84	65	90	2.762	113	93	99	3.618
25	1	3	1.212	54	20	41-43	1.960	83	64	89	2.736	112	92	99	3.591
24	1	2	1.185	53	18-19	39-40	1.934	82	62-63	89	2.709	111	92	98	3.564
15-23	1	1	0.[illegible] to 1.159	52	17	37-38	1.907	81	61	88	2.682	110	91	98	3.537

NH/djm
5-22-80

*Form D Norms for Florissant Valley were extrapolated from Form C, which were derived empirically.

NELSON-DENNY READING TEST NORMS
Grade 13, Beginning of Year - Form C
Prepared by
Neal E. Hartman
ST. LOUIS COMMUNITY COLLEGE AT FLORISSANT VALLEY

Total (V+C)	Natl. %ile	FV %ile	FV gpan	Total (V+C)	Natl. %ile	FV %ile	FV gpan	Total (V+C)	Natl. %ile	FV %ile	FV gpan	Total (V+C)	Natl. %ile	FV %ile	FV gpan
51	17	37-38	1.907	80	60-61	88	2.682	109	90	98	3.457	136			
50	15-16	35-36	1.880	79	58-59	87	2.655	108	89	98	3.431	135	99	99	4.000
49	14	32-33 34	1.864	78	57	86	2.629	107	88	98	3.404	134	98	99	4.000
48	13	30-31	1.827	77	55-56	84-85	2.602	106	88	97	3.377	133	98	99	4.000
47	13	28-29	1.800	76	54	83	2.575	105	87	97	3.350	132	98	99	4.000
46	12	26-27	1.773	75	52-53	82	2.548	104	86	97	3.324	131	98	99	4.000
45	11	24-25	1.747	74	51	81	2.523	103	85	97	3.297	130	98	99	4.000
44	10	22-23	1.720	73	49-50	80	2.495	102	84	97	3.270	129	98	99	3.991
43	9	20-21	1.693	72	47-48	79	2.468	101	83	97	3.243	128	97	99	3.965
42	8	19	1.666	71	46	77-78	2.442	100	83	97	3.217	127	97	99	3.938
41	7	17-18	1.640	70	44-45	75-76	2.415	99	82	96	3.190	126	97	99	3.912
40	7	16	1.613	69	43	73-74	2.388	98	81	96	3.163	125	97	99	3.885
39	6	14-15	1.586	68	41-42	71-72	2.361	97	80	96	3.137	124	97	99	3.858
38	6	13	1.560	67	40	70	2.335	96	79	96	3.110	123	97	99	3.831
37	5	11-12	1.533	66	38-39	68-69	2.308	95	78	95	3.083	122	96	99	3.805
36	5	10	1.506	65	36-37	66-67	2.281	94	77	95	3.056	121	96	99	3.778
35	4	9	1.479	64	35	64-65	2.254	93	76	95	3.030	120	96	99	3.751
34	4	8	1.453	63	33-34	63	2.228	92	75	94	3.003	119	95	99	3.725
33	3	76	1.426	62	32	61-62	2.201	91	74	94	2.976	118	95	99	3.698
32	3	7	1.399	61	30-31	59-60	2.174	90	73	94	2.949	117	94	99	3.671
31	2	6	1.372	60	29	56-58	2.148	89	72	93	2.923	116	93	99	3.644
30	2	5	1.346	59	28	54-55	2.121	88	71	92	2.896	115	93	99	3.618
29	2	4	1.319	58	27	52-53	2.094	87	70	92	2.869	114	92	99	3.591
28	1	4	1.292	57	25-26	50-51	2.067	86	69	92	2.843	113	92	98	3.564
27	1	3	1.266	56	24	48-49	2.041	85	67-68	91	2.816	112	91	98	3.537
26	1	3	1.239	55	22-23	46-47	2.014	84	66	90	2.789	111	90	98	3.511
25	1	3	1.212	54	21	44-45	1.987	83	65	90	2.762	110	90	98	3.484
24	1	2	1.185	53	19-20	41-43	1.960	82	63-64	89	2.736				
15-23	1	1	0.985 or 1.159	52	18	39-40	1.934	81	62	88	2.709				

NH/djm
5-22-80

ERIC Clearinghouse for Junior Colleges
8118 Math-Sciences Building
University of California
Los Angeles, California 90024

DEC 23 1983